APPROXIMATION BY STEP FUNCTIONS

GEORG AUMANN

1. Introduction. In a recent paper [**5**] I studied the Chebyshev approximation problem

$$f \sim \sum_{i=1}^{p} x_i \chi_i, \tag{1}$$

i.e., the approximation of a given bounded real function f on a set A by linear combinations of given characteristic functions $\chi_1, \cdots, \chi_p$ of subsets $A_1, \cdots, A_p$ of A in the sense of minimizing the norm

$$\|f - \sum x_i \chi_i\| = \sup\{ |f(a) - \sum x_i \chi_i(a)| : a \in A\}$$

by a proper choice of the x_i. As the problem is one of linear programming namely to find x_i and s such that

$$-s \leqq f(a) - \sum x_i \chi_i(a) \leqq s \qquad \text{for all } a \in A$$

and such that s is minimal, several methods to get a solution are at hand. Here we are concerned with a method which is especially adapted to the problem and which in case of the "matrix problem"

$$a_{ik} \sim x_i + y_k, \tag{2}$$

i.e., of approximating a given matrix (a_{ik}) by a matrix of the particular type $(x_i + y_k)$, has proved to be very efficient. It is the "leveling process" [**1**; **2**; **3**] which roughly speaking for problem (1) consists in an alternatively repeated minimizing within the sets A_i (in problem (2) the rows and columns). In [**4**] I pointed out by an example that the effectiveness of the leveling process depends on the structure of the covering of A by the A_i's and in [**5**] a decisive combinatorial property of the covering was introduced. The theorem which shows the bearing of this property on the approximation problem is here stated in the form of a necessary and sufficient condition. The examples given below disclose the surprising fact that approximation problems of the simple type as

$$a_{ijk} \sim x_i + y_j + z_k, \tag{3}$$

$$a_{ijk} \sim x_{ik} + y_{jk} + z_{ki}, \tag{4}$$

$$a_{jk} \sim x_j + y_k + z_{j+k}, \tag{5}$$

Received by the editors June 1961.

do not possess the property in question if the index sets are sufficiently large and therefore may be insensible to the leveling process.

2. For simplification we consider a finite set A and a covering $\Gamma=\{A_i: i\in I\}$ of A by a finite family of subsets A_i of A, $A=\bigcup_{i\in I} A_i$, $I=\{1,\cdots,p\}$. With this covering is associated the family Φ of all linear combinations

$$\phi = \sum x_i\chi_i$$

where χ_i denotes the characteristic function of A_i and the x_i's are real numbers. We consider the Chebyshev approximation of a given function $f_0|A$ by functions ϕ of Φ. If we define the norm $\|f\| =\max\{|f(a)|: a\in A\}$ for $f|A$ we have to find a $\phi_0\in\Phi$ such that

$$\|f_0-\phi_0\| \leqq \|f_0-\phi\| \text{ for all } \phi\in\Phi.$$

The leveling process presents itself if we reformulate the problem. Let us say that $f|A$ and $g|A$ are equivalent (with respect to the given covering Γ) if $f-g\in\Phi$ then our problem is this: Given a function f_0; find an equivalent one, say f^*, with least norm. For if f^* is equivalent to f_0 and of least norm we have $f_0=f^*+\phi_0$ with $\phi_0\in\Phi$ and $\phi_0=f_0-f^*$ is a solution of the approximation problem. So we have to work within the equivalence class of f_0 towards functions of smaller and smaller norms. The simplest way to produce a function equivalent to f is the transition $f\to f+y\chi_i$. If we take for y the value

$$y_0 = -\tfrac{1}{2}(\max f|A_i + \min f|A_i),$$

we have done the best for decreasing the norm. The transition

$$f\to f^{(i)} = f-\tfrac{1}{2}(\max f|A_i+\min f|A_i)\chi_i$$

is called the *leveling of f on A_i*. We evidently have $\|f^{(i)}\|\leqq\|f\|$. So leveling is a step towards a solution and so it seems quite natural to apply iterations of the leveling on the different A_i alternatively and in some periodic fashion.

3. We start with $f_0|A$, define

$$Lf = (\cdots((f^{(1)})^{(2)})\cdots)^{(p)}$$

and

$$f_n = L(f_{n-1}), \qquad n=1,2,\cdots.$$

Because of $\|f_{n+1}\|\leqq\|f_n\|$ we have the existence of $\lim_n\|f_n\|=b$. The surprising fact is that b may be larger than

$$\inf\{\|f_0-\phi\|: \phi\in\Phi\},$$

the minimal approximation error, a defect which eventually cannot be repaired by a rearrangement in the order of the different levelings. This is shown by the following *example* (there are simpler ones [4] but we use the one here for another reason):

Let A be the set $\{1, 2, \cdots, 8\}$ and A_i be

$$A_1' = \{1, 2\}, \quad A_2' = \{3, 4\}, \quad A_3' = \{5, 6\}, \quad A_4' = \{7, 8\};$$
$$A_1'' = \{1, 3\}, \quad A_2'' = \{4, 5\}, \quad A_3'' = \{2, 7\}, \quad A_4'' = \{6, 8\};$$
$$A_1''' = \{3, 5, 7\}, \quad A_2''' = \{1, 4, 8\}, \quad A_3''' = \{2, 6\}.$$

Consider the function f_0 with the values $f_0(1) = f_0(4) = f_0(6) = f_0(7) = 100$ and the value -100 on all other places. Then f_0 is already leveled on each A_i. So leveling is ineffective. But we can get a function $\bar{f}$ equivalent to f_0 with smaller norm by adding to each place on the sets $A_1', \cdots, A_3'''$ in the same order as listed above the values

$$31, 25, 19, 9; \qquad -24, -18, -10, 0; \qquad 0, -8, -10,$$

and for the resulting function $\bar{f}$ we find $\|\bar{f}\| = 99$.

4. The inefficiency of the leveling process depends on the structure of the covering. Because the convergence of the sequence f_n of §3 is a highly intricate matter—but knowing that if there is convergence the limit function is leveled on all A_i and equivalent to the original function—we ask an *intermediate question*: Under what conditions on the covering are we allowed to conclude that a function g, equivalent to f_0 and leveled on all A_i, i.e., $g^{(i)} = g$ for $i \in I$, is of least norm? To give an answer to this question we define:

A function $\sigma | A$ is said to be an *AS-function* ("function of alternating sign") with respect to the covering $\Gamma = \{A_i : i \in I\}$ if σ is not identically zero and

1. $\sigma(a) \in \{-1, 0, 1\}$ for all $a \in A$;
2. whenever $\sigma | A_i \neq 0$ there are at least two points x_1, x_2 on A_i with $\sigma(x_1) = 1$ and $\sigma(x_2) = -1$, $i \in I$.

A covering Γ is called an *L-covering* (the L simply indicates the reference to the leveling process) if to each AS-function $\sigma | A$ there is a function $s | A$ not identically zero and satisfying

(L′) sign $s(x) \in \{0, \sigma(x)\}$ for all $x \in A$;

(L″) $\sum_{x \in A_i} s(x) = 0$ for $i \in I$.

With these definitions we can state the

THEOREM. *If $\{A_i : i \in I\}$ is an L-covering of A then each function g equivalent to f and leveled on each A_i yields in $\phi = f - g$ a best Chebyshev approximation of f by linear combinations of the characteristic functions of the A_i's. And conversely, if this is true for any f then $\{A_i : i \in I\}$ is an L-covering.*

PROOF.

1. Let us assume that there is a function g leveled on all A_i and satisfying $\|g\| > \|f\|$ for some function f equivalent to g. We are going to show that $\Gamma = \{A_i : i \in I\}$ is no L-covering. Let $\|g\| = a$, then $a > 0$ and

$$\sigma(x) = \begin{cases} +1 & \text{if } g(x) = a, \\ -1 & \text{if } g(x) = -a, \\ 0 & \text{elsewhere} \end{cases}$$

defines an AS-function. With $f = g + \sum y_i\chi_i$ we get the inequalities

$$\sum y_i\chi_i(x) < 0 \text{ for } \sigma(x) = 1, \qquad \sum y_i\chi_i(x) > 0 \text{ for } \sigma(x) = -1.$$

Now assume that Γ is an L-covering. Then there is a function $s|A$ not identically zero and satisfying (L′) and (L″). This gives

$$\sum_{x\in A}\left(\sum_{i\in I} y_i\chi_i(x)\right)s(x) < 0.$$

The left side may be written

$$\sum_{i\in I} y_i \sum_{x\in A} \chi_i(x)s(x) = \sum_i y_i \sum_{x\in A_i} s(x) = 0.$$

This is a contradiction and the sufficiency of the condition is proved.

2. Now let us assume that for any g leveled on all A_i equivalent to f we have $\|g\| \leqq \|f\|$. Then for any AS-function $\sigma|A$ and any numbers y_i we have

$$\|\sigma + \sum y_i\chi_i\| \geqq 1. \tag{6}$$

Define $\sigma_i(x) = \sigma(x)\chi_i(x)$ and $A' = \{x : x \in A \text{ and } \sigma(x) \neq 0\}$. Then

(*) *not all the numbers* $\sum y_i\sigma_i(x)$, $x \in A'$, *are of the same sign.*

For if for instance all these numbers would be <0 then with some $\rho > 0$ we could replace the y_i by $y_i' = \rho y_i$ and arrive at $\|\sigma + \sum y_i'\chi_i\| < 1$ in contradiction to (6). But (*) is a well-known [6] sufficient condition that the system

$$\sum_{x\in A} \sigma_i(x)S(x) = 0, \qquad S \geqq 0$$

allows a solution S not identically zero. With $s(x) = \sigma(x)S(x)$ we see that (L′), (L″) can be satisfied with $s \neq 0$.

5. Examples of L-coverings.

PROPOSITION 1. *For every $n \times m$-matrix the system of rows and columns is an L-covering.*

PROOF. Every AS-function σ on the matrix array contains an irreducible AS-function σ' which on a row or column where it is not zero yields exactly one $+1$ and one -1. Evidently σ' is a function s fitting to σ in the sense of the theorem above.

PROPOSITION 2. *If* $\{A_i: i\in I\}$ *is an L-covering of* A *and* $B\subset A$ *then the "trace covering" on* B, $\{B\cap A_i: i\in A$ *and* $B\cap A_i\neq\emptyset\}$ *is an L-covering of* B.

PROOF. Let σ be an AS-function with respect to the trace covering. We extend it by defining $\sigma|(A-B)=0$ and get an AS-function with respect to $\{A_i: i\in I\}$. By our theorem we have a function $s|A$ fitting to $\sigma|A$. Evidently $s|B$ fits to $\sigma|B$ with respect to the trace covering.

6. As a matter of fact there are many simple coverings which are no L-coverings.

The coverings belonging to the approximation problems (3), (4) *and* (5) *are no L-coverings if the index sets are large enough.*

PROOF.

Concerning (3). The covering of the example in §3 can be considered as a trace covering on a cubic $4\times 4\times 3$-matrix covered by its 2-dimensional layers. So by Proposition 2 a cubic matrix with at least four 2-dimensional layers in each direction is no L-covering.[1]

Concerning (4). Consider the following AS-function σ on a cubic $5\times 5\times 6$-matrix

1+ 2−	4+ 5−	1− 3+	4− 2+	5+ 3−
6− 2+	3− 5+	6+ 5−	6+ 2−	3+ 2−
3− 6+	3+ 4−	3+ 6−	6− 4+	
		1+ 3−	3+ 2−	1− 2+
1− 3+	4− 5+	1− 5+	3− 4+	1+ 5−

[1] This fact disproves a hypothesis of M. Golomb [3, p. 324, (10.53)].

where the figure indicates the height of the layer and the sign behind it the sign of σ. On all other places σ has the value 0. It is easy to check that any function s satisfying (L′) and (L″) with respect to σ is identically zero. So the covering of problem (4) is no L-covering.

Concerning (5). We use the preceding example. We project its cubic matrix array into a (j, k)-plane in such a way that the rods of the matrix are projected into the lines $j=\text{const.}$, $k=\text{const.}$, and $j+k=\text{const.}$ and that no two rods have colinear images. So we see that the covering of the problem (4) is a trace covering of problem (5). Proposition 2 again proves that the covering of (5) is no L-covering as soon as there are sufficiently many layers in each family.

References

1. S. P. Diliberto and E. G. Strauss, *Approximation of a function of several variables by a sum of functions*, Pacific J. Math. **1** (1951), 195–210.

2. G. Aumann, *Über approximative Nomographie.* I, S.-B. Ber. Bayer. Akad. Wiss. Math. Nat. Kl. (1958), 137–155.

3. M. Golomb, *Approximations by functions of fewer variables*, Proc. Math. Research Center, U. S. Army, Univ. of Wisconsin, Publ. No. 1, *On Numerical Approximation*, Madison, Wisconsin, 1959.

4. G. Aumann, *Lineare Approximationen auf einem Geflecht*, Arch. Math. **10** (1959), 267–272.

5. ———, *Über approximative Nomographie.* III, S.-B. Ber. Bayer. Akad. Wiss. Math. Nat. Kl. (1960), 27–34.

6. H. W. Kuhn and A. W. Tucker, *Linear inequalities and related systems*, Ann. of Math. Studies Vol. 38, Princeton Univ. Press, Princeton, N. J., 1956, p. 8.

UNIVERSITY OF IDAHO AND
TECHNISCHE HOCHSCHULE MUNICH